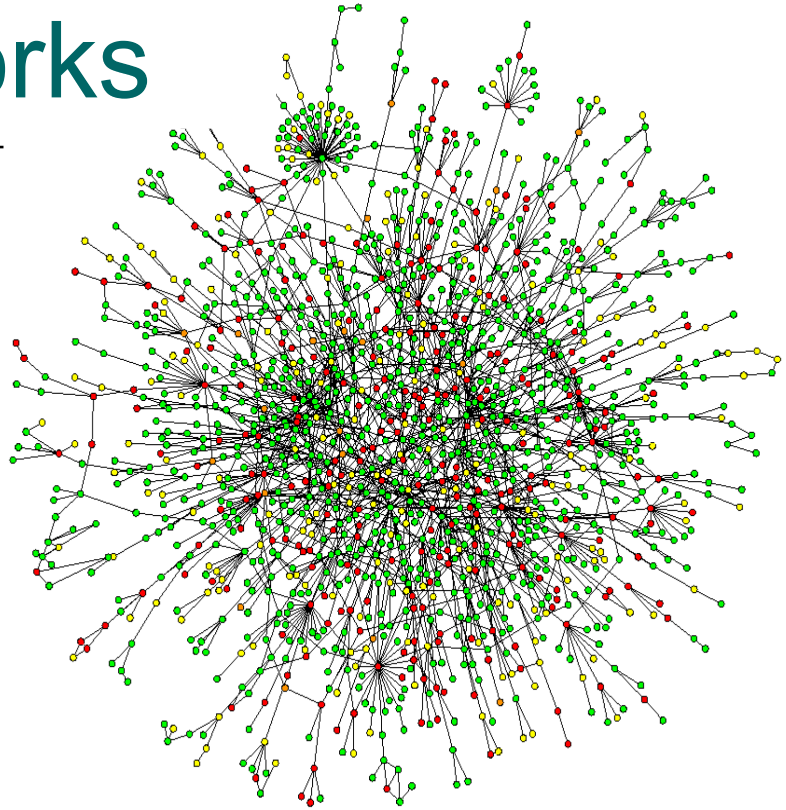
Scale Free Networks



Franco Zambonelli

February 2010

Outline

- Characteristics of Modern Networks
  - Small World & Clustering
  - Power law Distribution
  - Ubiquity of the Power Law
- Deriving the Power Law
  - How does network grow?
  - The theory of preferential attachment
  - Variations on the theme
- Properties of Scale Free Networks
  - Error, attack tolerance, and epidemics
  - Implications for modern distributed systems
  - Implications for everyday systems
- Conclusions and Open Issue

Part 1

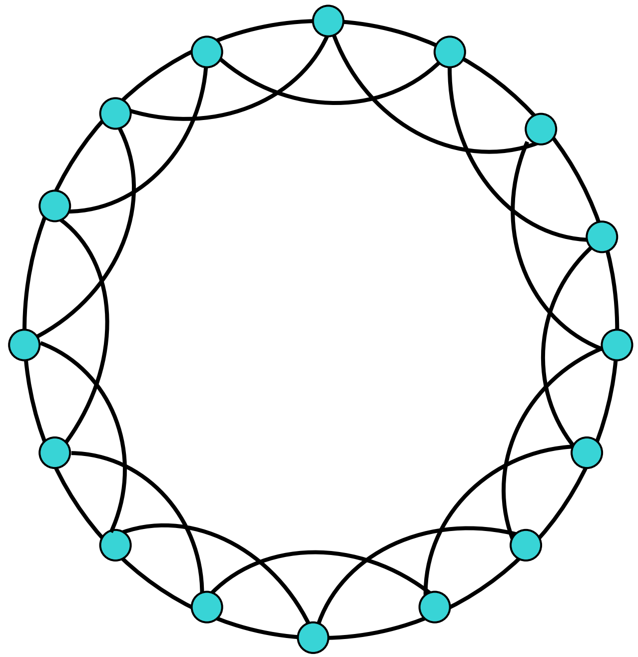
- Characteristics of Modern Networks

Characteristics of Modern Networks

- Most networks
  - Social
  - Technological
  - Ecological
- Are characterized by being
  - Small world
  - Clustered
  - And SCALE FREE (Power law distribution)
- We now have to understand
  - What is the power law distribution
  - And how we can model it in networks

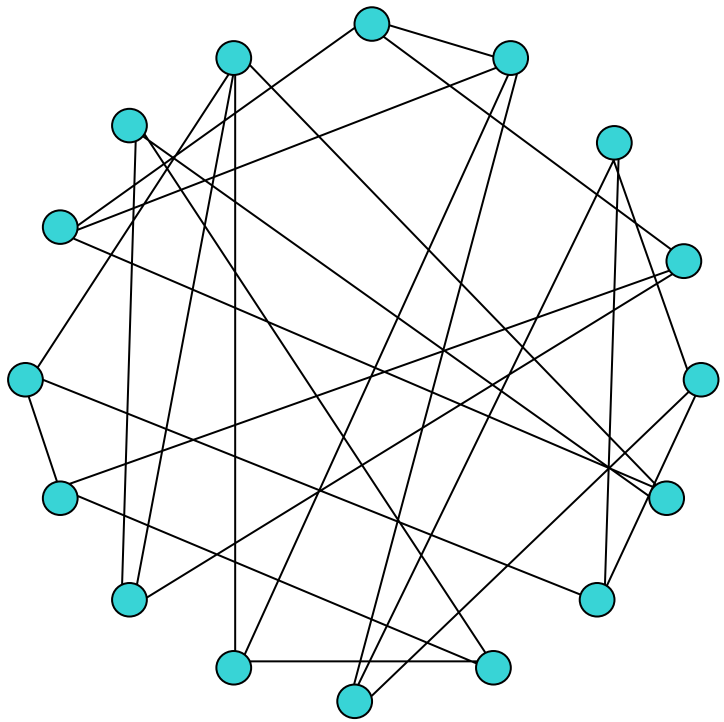
Regular Lattice Networks

- Nodes are connected in a regular neighborhood
  - They are usually k-regular, with a fixed number k of edges per each node
- They do not exhibit the small world characteristics
  - The average distance between nodes grown with the d-root of n, where n is the number of nodes
- They do may exhibit clustering
  - Depending on the lattice and on the k factor, neighbor nodes are also somehow connected with each other



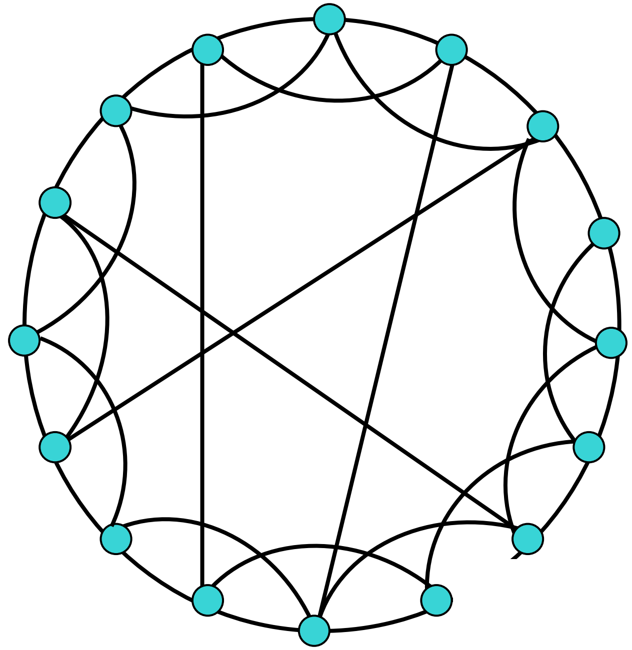
Random Networks

- Random networks have randomly connected edges
  - If the number of edges is M, each node has an average of k=M/2n edges, where n is the number of nodes
- They exhibit the small world characteristics
  - The average distance between nodes is log(n), where n is the number of nodes
- They do not exhibit clustering
  - The clustering factor is about C=k/n for large n



Small World Networks

- Watts and Strogatz (1999) propose a model for networks “between order and chaos”
- Such that
  - The network exhibit the small world characteristic, as random networks
  - And at the same exhibit relevant clustering, as regular lattices
- The model is built by simply
  - Re-wiring at random a small percentage of the regular edges
  - This is enough to dramatically shorten the average path length, without destroying clustering



The Degree Distribution

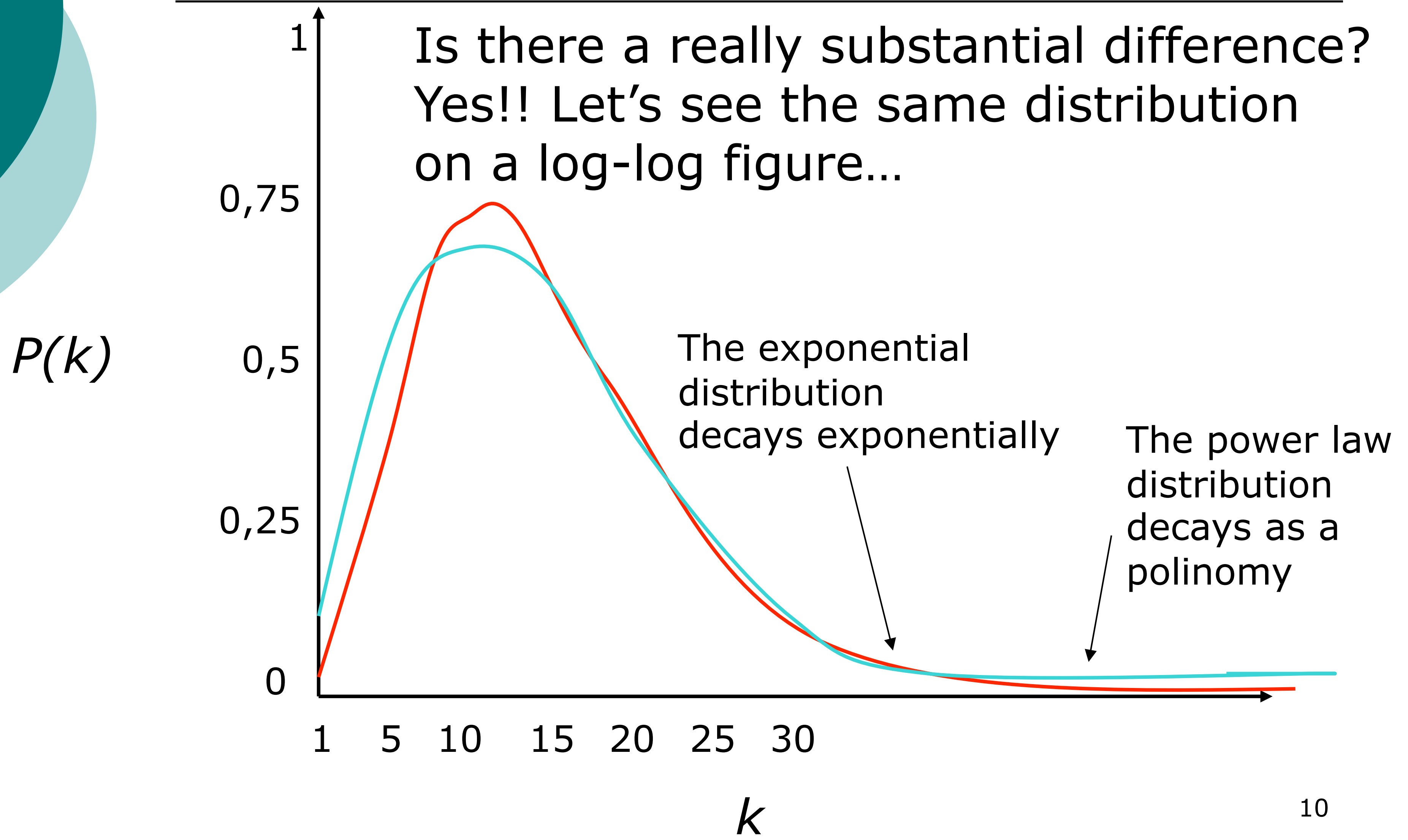
- What is the degree distribution?
  - It is the way the various edges of the network "distributes" across the vertices
  - How many edges connect the various vertices of the network
- For the previous types of networks
- In k-regular regular lattices, the distribution degree is constant
  - $P(k_r)=1$ for all nodes (all nodes have the same fixed k_r number of edges)
- In random networks, the distribution can be either constant or exponential
  - $P(k_r)=1$ for all nodes (is the randon network has been constructed as a k-regular network)
  - $P(k_r)=\alpha e^{-\beta k}$, that is the normal "gaussian" distribution, as derived from the fact that edges are independently added at random

The Power Law Distribution

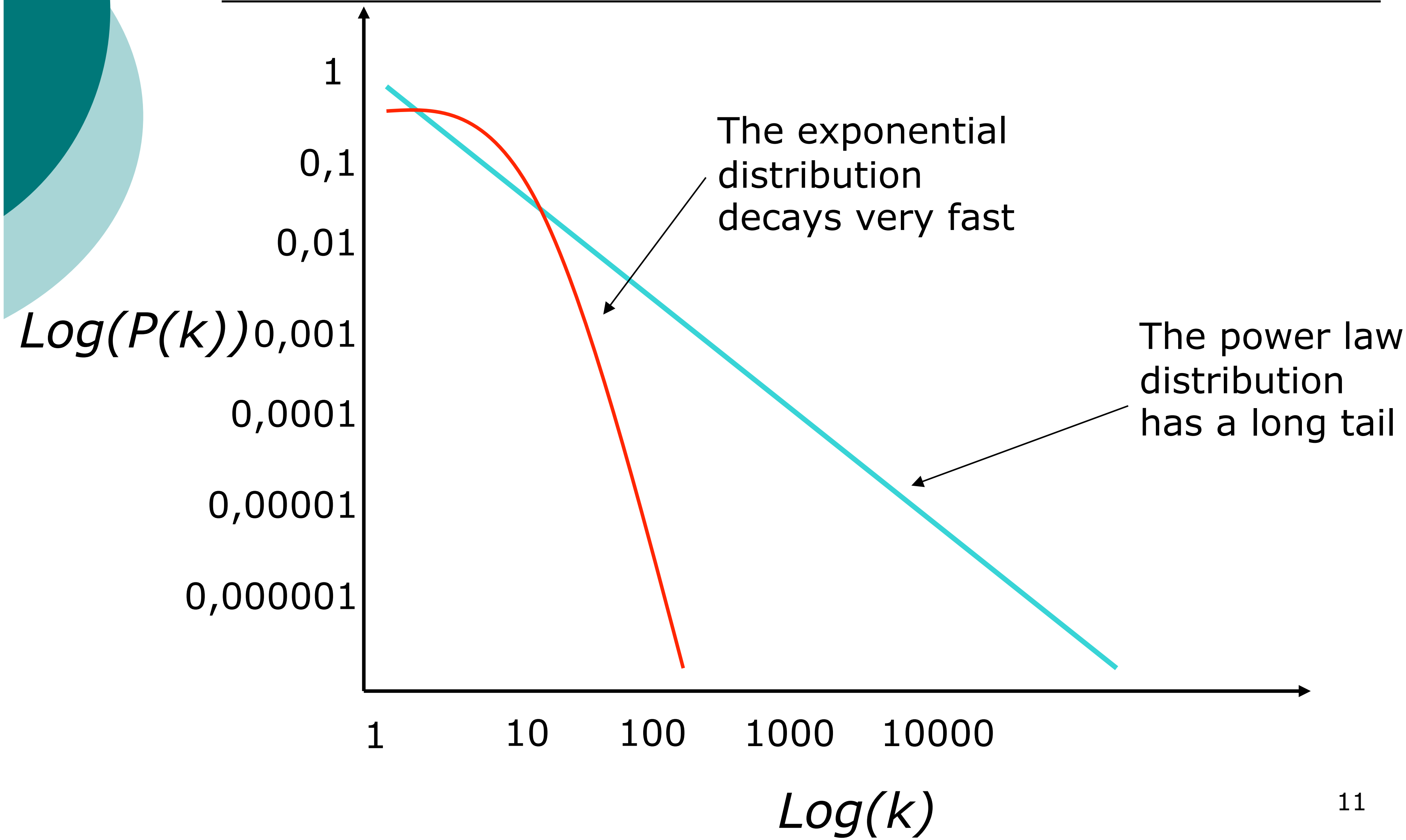
- Most real networks, instead, follow a “power law” distribution for the node connectivity
- In general term, a probability distribution is “power law” if
  - The probability *P(k)* that a given variable *k* has a specific value
  - Decreases proportionally to *k* power $-\gamma$, where γ is a constant value
- For networks, this implies that
  - The probability for a node to have k edges connected
  - Is proportional to $\alpha k^{-\gamma}$

$$P(k) = \alpha k^{-\gamma}$$

Power vs. Exponential Distribution



Power vs. Exponential Distribution



The Heavy Tail

- The power law distribution implies an "infinite variance"
  - The "area" of "big *ks*" in an exponential distribution tend to zero with $k \rightarrow \infty$
  - This is not true for the power law distribution, implying an infinite variance
  - The tail of the distribution counts!!!
- In other words, the power law implies that
  - The probability to have elements very far from the average is not neglectable
  - The big number counts
- Using an exponential distribution
  - The probability for a Web page to have more than 100 incoming links, considering the average number of links for page, would be less in the order of 1^{-20}
  - which contradicts the fact that we know a lot of "well linked" sites…

The Power Law in Real Networks

Average k

Power law exponents

Network	Size	$\langle k \rangle$	κ	γ_{out}	γ_{in}	ℓ_{real}	ℓ_{rand}	ℓ_{pow}	Reference
WWW	325 729	4.51	900	2.45	2.1	11.2	8.32	4.77	Albert, Jeong, and Barabási 1999
WWW	4×10^7	7		2.38	2.1				Kumar *et al.*, 1999
WWW	2×10^8	7.5	4000	2.72	2.1	16	8.85	7.61	Broder *et al.*, 2000
WWW, site	260 000				1.94				Huberman and Adamic, 2000
Internet, domain*	3015–4389	3.42–3.76	30–4	2.1–2.2	2.1–2.	4	6.3	5.2	Faloutsos, 1999
Internet, router*	3888	2.57	30	2.48	2.48	12.15	8.75	7.67	Faloutsos, 1999
Internet, router*	150 000	2.66	60	2.4	2.4	11	12.8	7.47	Govindan, 2000
Movie actors*	212 250	28.78	900	2.3	2.3	4.54	3.65	4.01	Barabási and Albert, 1999
Co-authors, SPIRES*	56 627	173	1100	1.2	1.2	4	2.12	1.95	Newman, 2001b
Co-authors, neuro.*	209 293	11.54	400	2.1	2.1	6	5.01	3.86	Barabási *et al.*, 2001
Co-authors, math.*	70 975	3.9	120	2.5	2.5	9.5	8.2	6.53	Barabási *et al.*, 2001
Sexual contacts*	2810			3.4	3.4				Liljeros *et al.*, 2001
Metabolic, *E. coli*	778	7.4	110	2.2	2.2	3.2	3.32	2.89	Jeong *et al.*, 2000
Protein, *S. cerev.**	1870	2.39		2.4	2.4				Jeong, Mason, *et al.*, 2001
Ythan estuary*	134	8.7	35	1.05	1.05	2.43	2.26	1.71	Montoya and Solé, 2000
Silwood Park*	154	4.75	27	1.13	1.13	3.4	3.23	2	Montoya and Solé, 2000
Citation	783 339	8.57			3				Redner, 1998
Phone call	53×10^6	3.16		2.1	2.1				Aiello *et al.*, 2000
Words, co-occurrence*	460 902	70.13		2.7	2.7				Ferrer i Cancho and Solé, 2001
Words, synonyms*	22 311	13.48		2.8	2.8				Yook *et al.*, 2001b

The Ubiquity of the Power Law

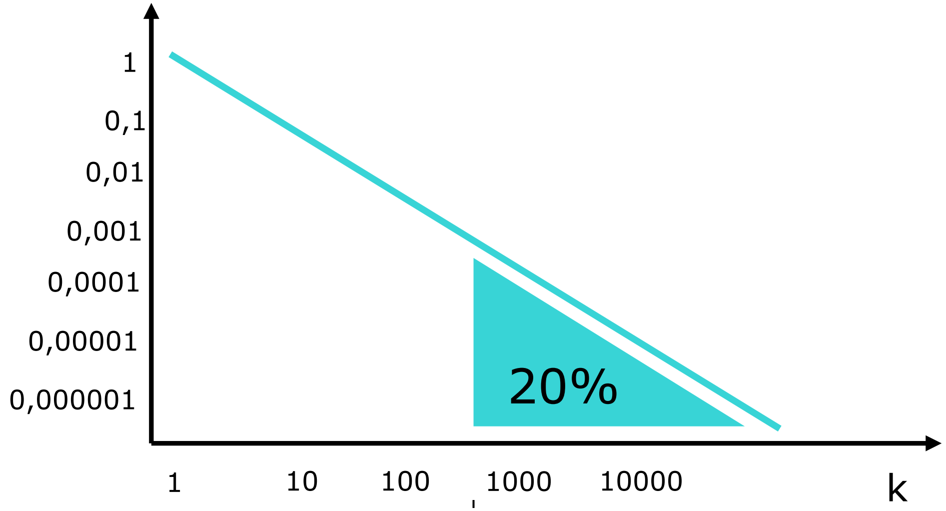
- The previous table include not only technological networks
  - Most real systems and events have a probability distribution that
  - Does not follow the “normal” distribution
  - And obeys to a power law distribution
- Examples, in addition to technological and social networks
  - The distribution of size of files in file systems
  - The distribution of network latency in the Internet
  - The networks of protein interactions (a few protein exists that interact with a large number of other proteins)
  - The power of earthquakes: statistical data tell us that the power of earthquakes follow a power-law distribution
  - The size of rivers: the size of rivers in the world is is power law
  - The size of industries, i.e., their overall income
  - The richness of people
  - In these examples, the exponent of the power law distribution is always around 2.5
- The power law distribution is the “normal” distribution for complex systems (i.e., systems of interacting autonomous components)
  - We see later how it can be derived…

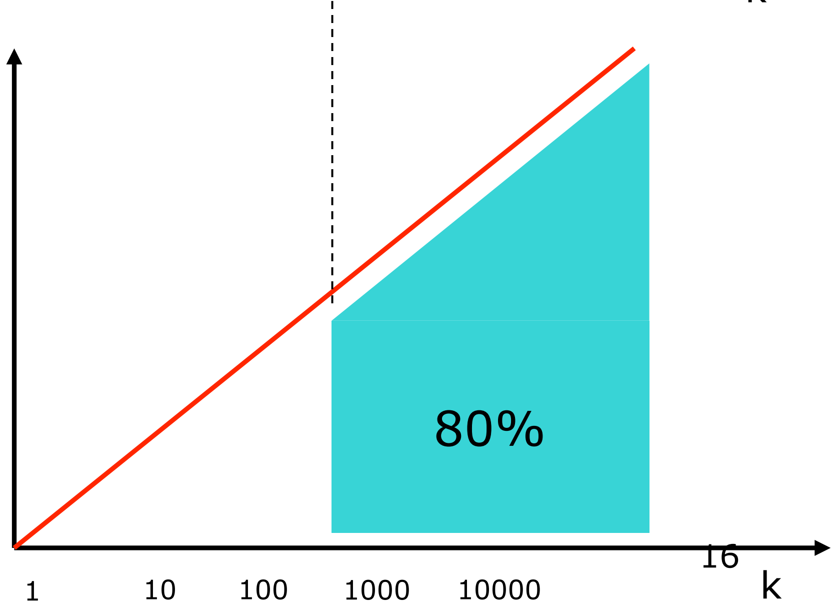
The 20-80 Rule

- It’s a common “way of saying”
  - But it has scientific foundations
  - For all those systems that follow a power law distribution
- Examples
  - The 20% of the Web sites gests the 80% of the visits (actual data: 15%-85%)
  - The 20% of the Internet routers handles the 80% of the total Internet traffic
  - The 20% of world industries hold the 80% of the world’s income
  - The 20% of the world population consumes the 80% of the world’s resources
  - The 20% of the Italian population holds the 80% of the lands (that was true before the Mussolini fascist regime, when lands re-distribution occurred)
  - The 20% of the earthquakes caused the 80% of the victims
  - The 20% of the rivers in the world carry the 80% of the total sweet water
  - The 20% of the proteins handles the 80% of the most critical metabolic processes
- Does this derive from the power law distribution? YES!

The 20-80 Rule Unfolded

- The 20% of the population
  - Remember the area represents the amount of population in the distribution
- Get the 80% of the resources
  - In fact, it can be found that the "amount of resources" (i.e., the amount of links in the network) is the integral of *P(k)*k*, which is nearly linear
- I know you have paid attention and would say the "25-75" rule, but remember there are bold approximations…



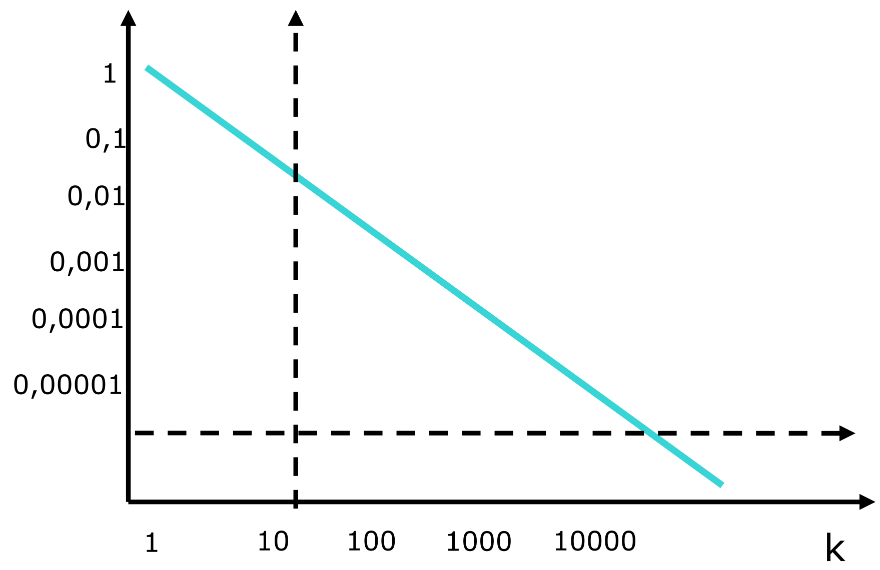


Hubs and Connectors

- Scale free networks exhibit the presence of nodes that
  - Act as hubs, i.e., as point to which most of the other nodes connects to
  - Act as connectors, i.e., nodes that make a great contributions in getting great portion of the network together
  - “smaller nodes” exists that act as hubs or connectors for local portion of the network
- This may have notable implications, as detailed below

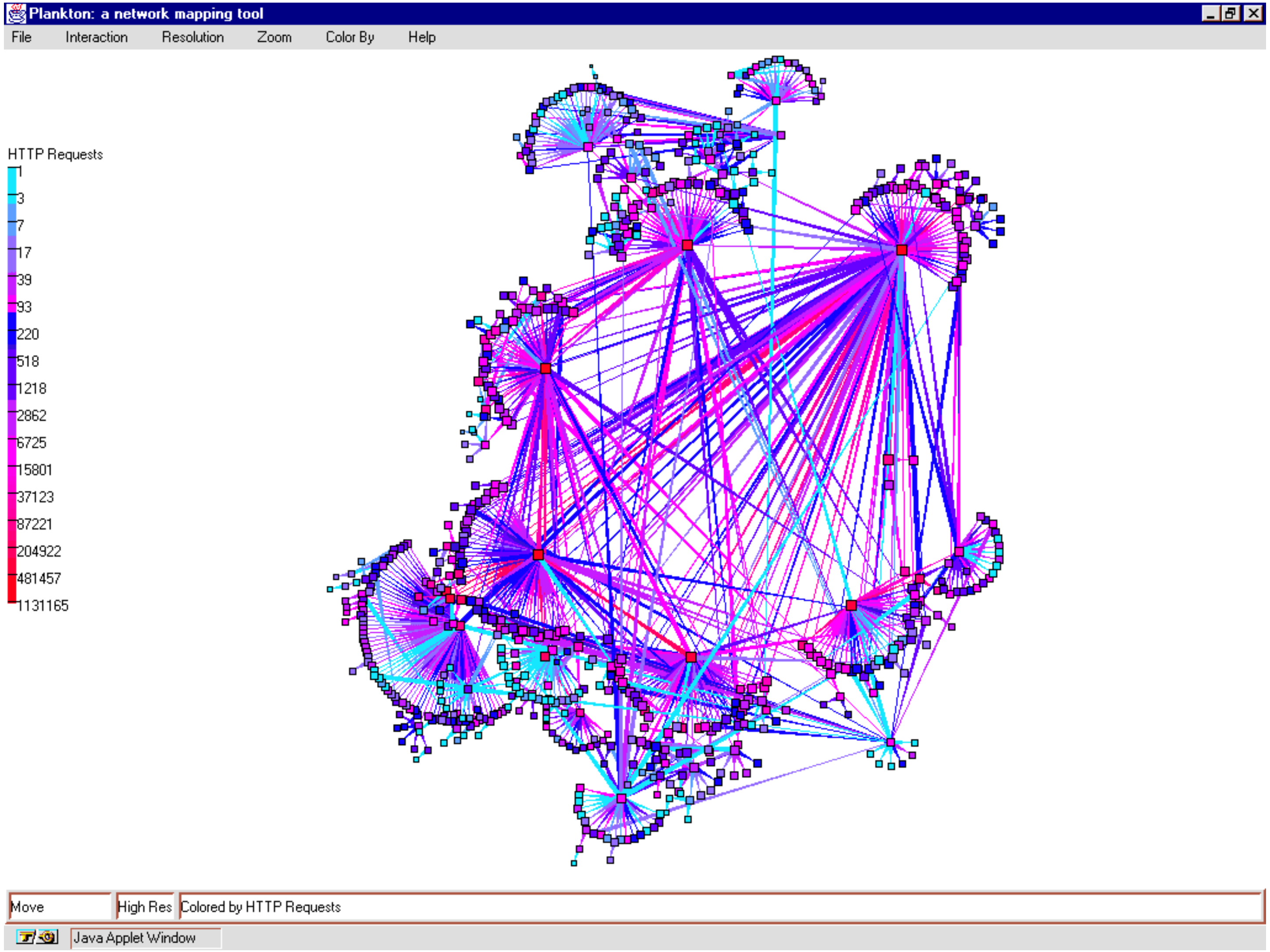
Why "Scale-Free" Networks

- Why networks following a power law distribution for links are called "**scale free**"?
  - Whatever the scale at which we observe the network
  - The network looks the same, i.e., it looks similar to itself
- The overall properties of the network are preserved independently of the scale
- In particular:
  - If we cut off the details of a network – skipping all nodes with a limited number of links – the network will preserve its power-law structure
  - If we consider a sub-portion of any network, it will have the same overall structure of the whole network



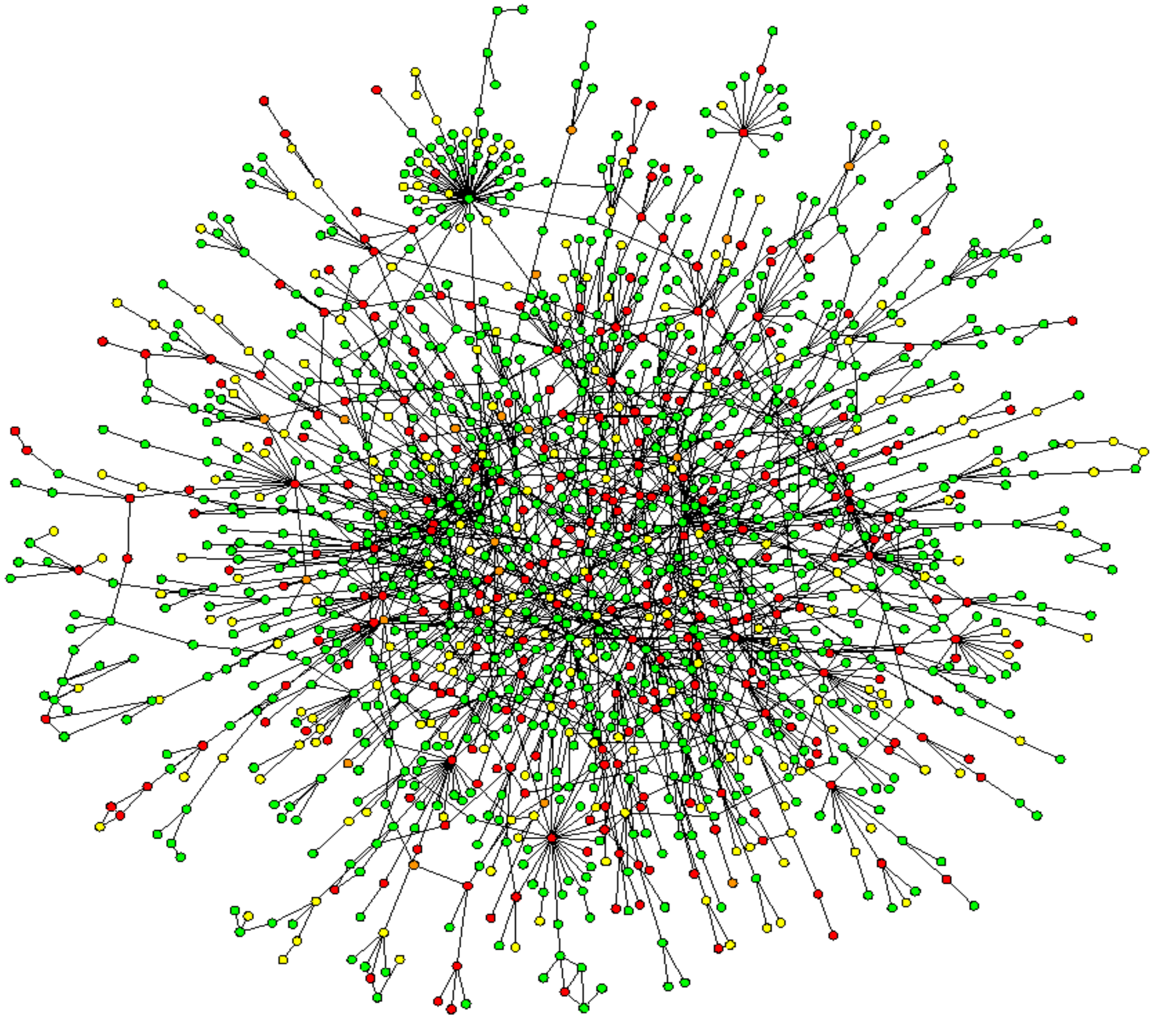
How do Scale Free Networks Look Like?

Web Cache Network



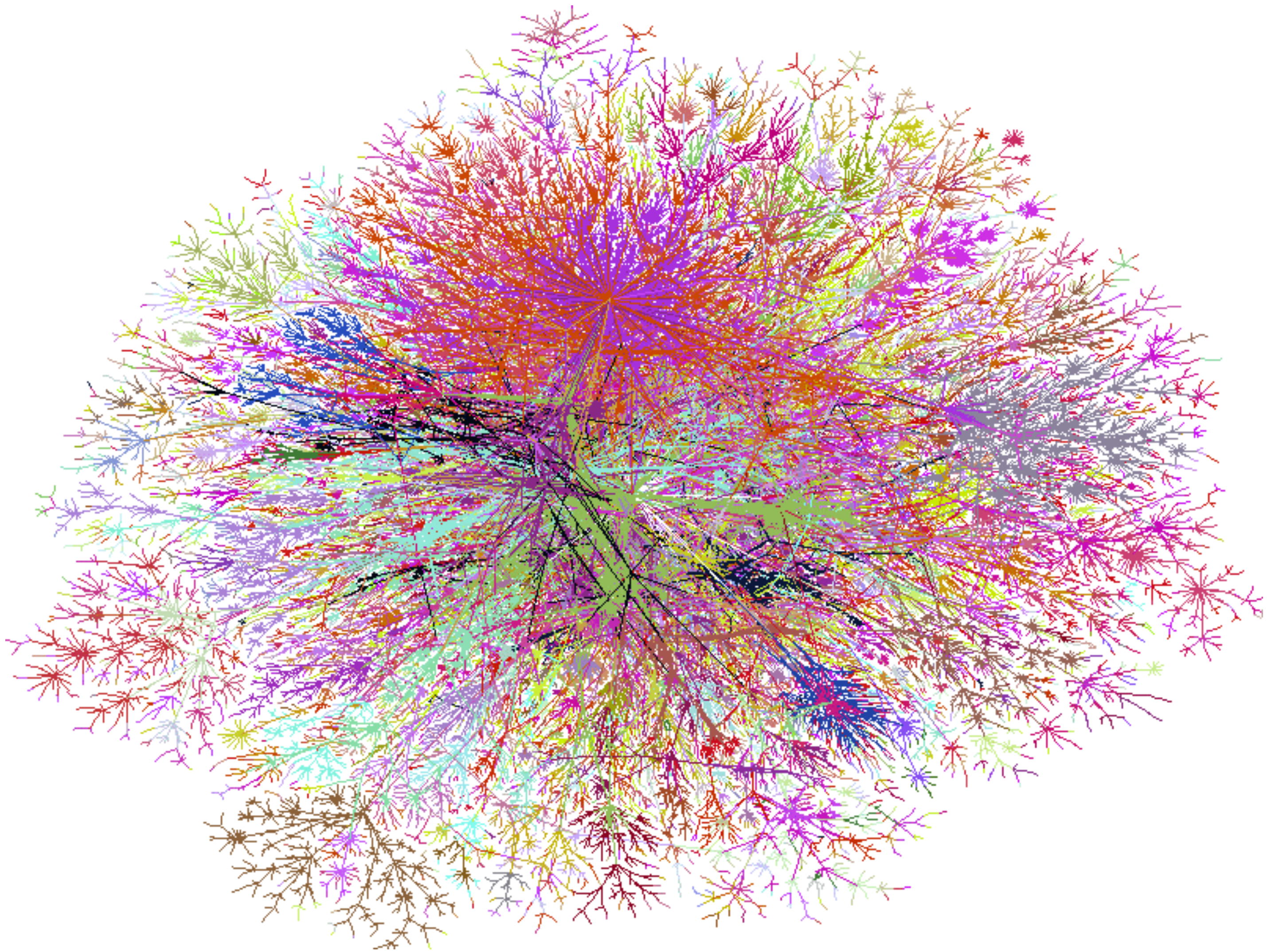
How do Scale Free Networks Look Like?

Protein Network



How do Scale Free Networks Look Like?

The Internet
Routers

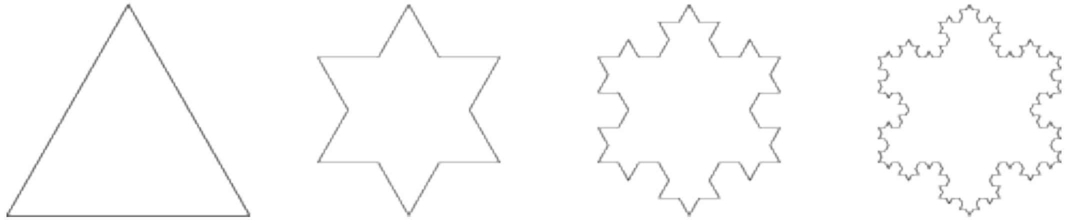


Fractals and Scale Free Networks

- The nature is made up of mostly “fractal objects”
- The fractal term derives from the fact that they have a non-integer dimension
  - 2-d objects have a “size” (i.e., a surface) that scales with the square of the linear size $A=kL^2$
  - 3-d objects have a “size” (i.e., a volume) that scales with the cube of the linear size $V=kL^3$
  - Fractal objects have a “size” that scales with some fractions of the linear size $S=kL^{a/b}$
- Fractal objects have the property of being “self-similar” or “scale-free”
  - Their “appearance” is independent from the scale of observation
  - They are similar to itself independently of whether you look at the from near and from far
  - That is, they are scale-free

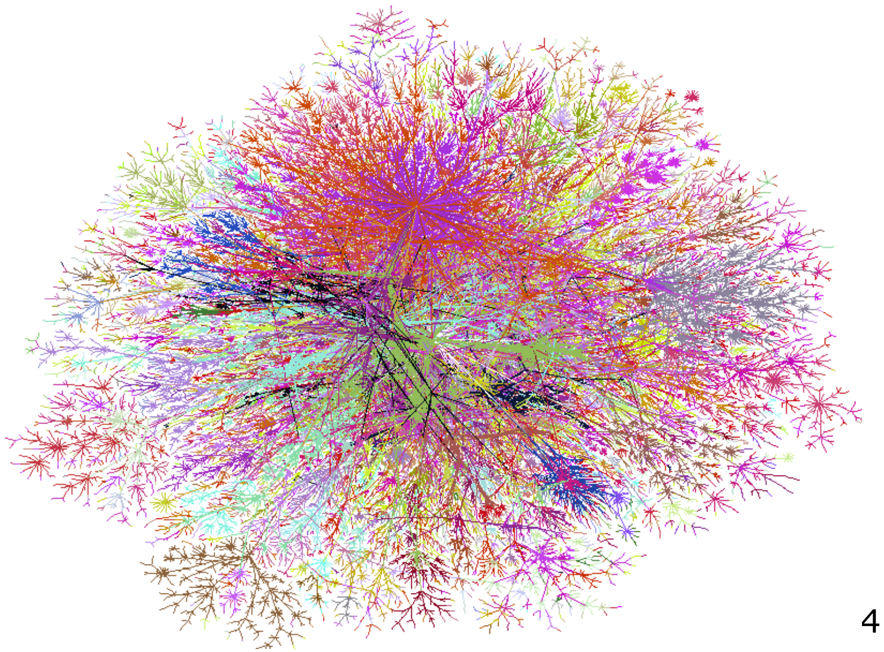
Examples of Fractals

- The Koch snowflake
- Coastal Regions & River systems
- Lymphatic systems
- The distribution of masses in the universe



Scale Free Networks are Fractals?

- Yes, in fact:
  - They are the same at whatever dimension we observe them
  - Also, the fact that they grow according to a power law can be considered as a sort of fractal dimension of the network…
- Having a look at the figures clarifies the analogy



Part 2

- Explaining the Power Law

Growing Networks

- In general, network are not static entities
- They grow, with the continuous addition of new nodes
  - The Web, the Internet, acquaintances, the scientific literature, etc.
  - Thus, edges are added in a network with time
- The probability that a new node connect to another existing node may depend on the characteristics of the existing node
  - This is not simply a random process of independent node additions
  - But there could be “preferences” in adding an edge to a node
  - E.g.,. Google, a well known and reliable Internet router, a cool guy who knows many girls, a famous scientist,
  - Both of these could attract more link…

Evolving Networks

- More in general…
  - Networks grows AND
  - Network evolves
- The evolution may be driven by various forces
  - Connection age
  - Connection satisfaction
- What matters is that connections can change during the life of the network
  - Not necessarily in a random way
  - But following characteristics of the network…
- Let's start with the growing process..

Preferential Attachment

- Barabasi and Albert shows that
- Making a network grow with new nodes that
  - Enter the network in successive times
  - Attach preferentially to nodes that already have many links
- Lead to a network structure that is
  - Small world
  - Sometimes Clustered
  - And Power-law: the distribution of link on the network nodes obeys to the power law distribution!

- Let’s call this the “BA model”

The Preferential Attachment Algorithm

- Start with a limited number of initial nodes
- At each time step, add a new node that has *m* edges that link to m existing nodes in the system
- When choosing the nodes to which to attach, assume a probability ∏ for a node i proportional to the number k_i of links already attached to it
- After t time steps, the network will have $n=t+m_0$ nodes and $M=mt$ edges
- **It can be shown that this leads to a power law network!**

$$m_0$$

$$m \le m_0$$

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

$$n = t + m_0$$

$$M = mt$$

Proof (1)

- Assume for simplicity that k_i for any node i is a continuous variable
- Because of the assumptions, k_i is expected to grow proportionally to $\Pi(k_i)$, that is to its probability of having a new edge
- Consequently, and because m edges are attached at each time, k_i should obey the differential equation aside

$$\frac{\partial k_i}{\partial t} = m\Pi(k_i) = m\frac{k_i}{\sum_{j=1}^{n-1} k_j}$$

Proof (2)

- The sum:

$$\sum_{j=1}^{n-1} k_j$$

- Goes over all nodes except the new ones
- This it results in:
- Remember that the total number of edges is *mt* and that here is edge is counted twice

$$\sum_{j=1}^{n-1} k_j = 2mt - m$$

- Substituting in the differential equation

$$\frac{\partial k_i}{\partial t} = m \frac{k_i}{\sum_{j=1}^{n-1} k_j} = m \frac{k_i}{2mt - m} \approx m \frac{k_i}{2t}$$

Proof (3)

- We have now to solve this equation:
  - That is, we have to find a $k_i(t)$ function such as its derivative is equal to: itself, mutiplied by m, and divided by $2t$
- We now show this is:
- In fact:

$$\frac{\partial k_i}{\partial t} = m \frac{k_i}{2t}$$

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta}; \quad \text{with } \beta = \frac{1}{2}$$

$$\frac{\partial}{\partial t}\left(m\left(\frac{t}{t_i}\right)^{\beta}\right) = \frac{1}{2}\frac{m}{t_i^{\beta}}\frac{1}{t^{\beta}} = \frac{1}{2}\frac{m}{t_i^{\beta}}\frac{1}{t^{\beta}}\frac{t^{\beta}}{t^{\beta}} = \frac{m}{2}\frac{t^{\beta}}{t_i^{\beta}}\frac{1}{t^{2\beta}} = \frac{k_i(t)}{2t}$$

- Where we also consider the initial condition $k_i(t_i)=m$, where t_i is the time at which node i has arrived

Proof (4)

- The $k_i(t)$ function that we have not calculated shows that the degree of each node grown with a power law with time
- Now, let's calculate the probability that a node has a degree $k_i(t)$ smaller than k
- *We have:*

$$P[k_i(t) < k] = P\left[m\frac{t^{\beta}}{t_i^{\beta}} < k\right] = P\left[m^{\frac{1}{\beta}}\frac{t^{\beta\frac{1}{\beta}}}{t_i^{\beta\frac{1}{\beta}}} < k^{\frac{1}{\beta}}\right] =$$

$$= P\left[m^{\frac{1}{\beta}}\frac{t}{t_i} < k^{\frac{1}{\beta}}\right] = P\left[t_i > \frac{m^{\frac{1}{\beta}}t}{k^{\frac{1}{\beta}}}\right]$$

Proof (5)

- Now let's remember that we add nodes at each time interval
- Therefore, the probability t_i for a node, that is the probability for a node to have arrived at time t_i is a constant and is:

$$P(t_i) = \frac{1}{t + m_0}$$

- Substituting this into the previous probability distribution:

$$P[k_i(t) < k] = P\left[t_i > \frac{m^{\frac{1}{\beta}} t}{k^{\frac{1}{\beta}}}\right] = 1 - P\left[t_i \le \frac{m^{\frac{1}{\beta}} t}{k^{\frac{1}{\beta}}}\right] = 1 - \frac{m^{\frac{1}{\beta}} t}{k^{\frac{1}{\beta}}(t + m_0)}$$

Proof (6)

- Now given the probability distribution:
- Which represents the probability that a node *i* has less than *k* link
- The probability that a node has exactly *k* link can be derived by the derivative of the probability distribution

$$P[k_i(t) < k]$$

$$P(k) = \frac{\partial P[k_i(t) < k]}{\partial k}$$

$$P(k) = \frac{\partial P[k_i(t) < k]}{\partial k} = \frac{\partial}{\partial k}(1 - \frac{m^{\frac{1}{\beta}} t}{k^{\frac{1}{\beta}} (t + m_0)}) = \frac{2m^{\frac{1}{\beta}} t}{m_0 + t} \frac{1}{k^{\frac{1}{\beta}+1}}$$

Conclusion of the Proof

- Given *P(k)*:

$$P(k) = \frac{2m^{\frac{1}{\beta}} t}{m_0 + t} \frac{1}{k^{\frac{1}{\beta}+1}}$$

- After a while, that is for t→∞

$$P(k) \approx 2m^{\frac{1}{\beta}} k^{-\frac{1}{\beta}-1} = 2m^{\frac{1}{\beta}} k^{-\gamma} \quad \text{where } \gamma = \frac{1}{\beta} + 1 = 3$$

- That is, **we have obtained a power law probability density**, with an exponent which is independent of any parameter (being the only initial parameter m)

Probability Density for a Random Network

- In a random network model, each new node that attach to the network attach its edges independently of the current situation
  - Thus, all the events are independent
- The probability for a node to have a certain number of edges attached is thus a “normal”, exponential, distribution
- It can be easily found, using standard statistical methods that:

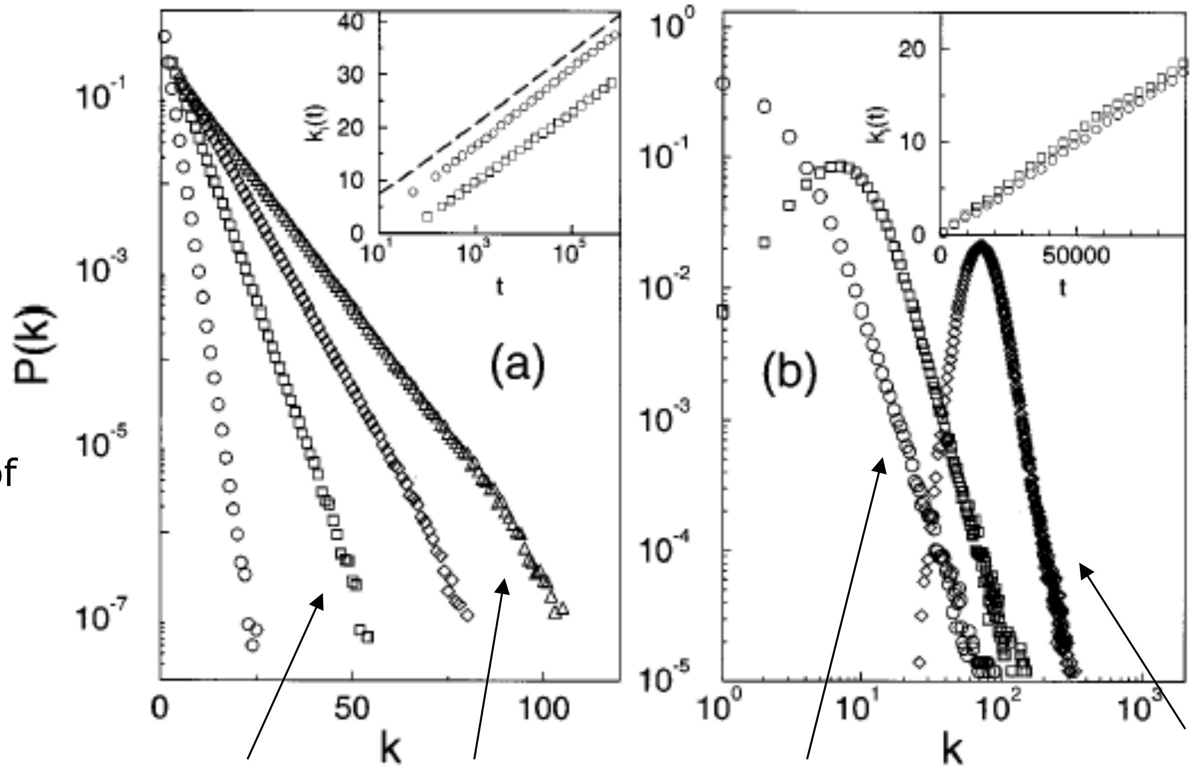
$$P(k) = \frac{1}{m} e^{-\frac{k}{m}}$$

Barabasi-Albert Model vs. Random Network Model

- See the difference for the evolution of the Barabasi-Albert model vs. the Random Network mode (from Barabasi and Albert 2002)

Barabasi-Albert Model n=800000

Simulations performed with various values of m



Random network model for n=10000

The degree distribution gradually becomes a normal one with passing time

m=3

m=7

t=n

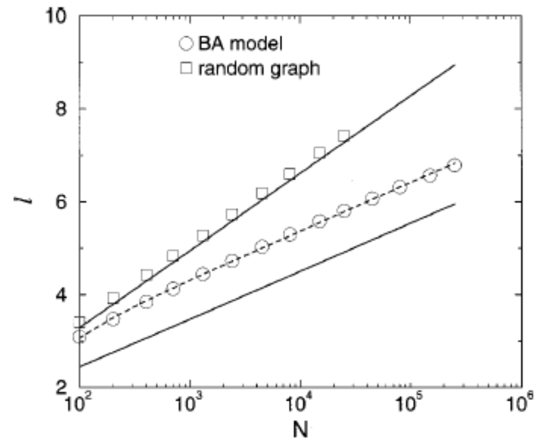
t=50n

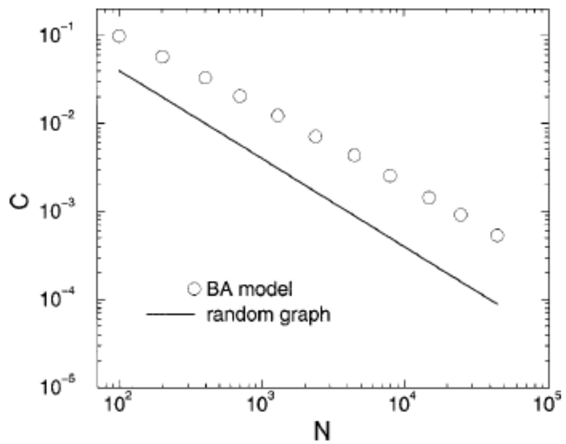
Generality of the Barabasi-Albert Model

- In its simplicity, the BA model captures the essential characteristics of a number of phenomena
  - In which events determining "size" of the individuals in a network
  - Are not independent from each other
  - Leading to a power law distribution
- So, it can somewhat explain why the power law distribution is as ubiquitous as the normal Gaussian distribution
- Examples
  - **Gnutella**: a peer which has been there for a long time, has already collected a strong list of acquaintances, so that any new node has higher probability of getting aware of it
  - **Rivers**: the eldest and biggest a river, the more it has probability to break the path of a new river and get its water, thus becoming even bigger
  - **Industries**: the biggest an industry, the more its capability to attract clients and thus become even bigger
  - **Earthquakes**: big stresses in the earth plaques can absorb the effects of small earthquakes, this increasing the stress further. A stress that will eventually end up in a dramatic earthquakes
  - **Richness**: the rich I am, the more I can exploit my money to make new money → "RICH GET RICHER"

Additional Properties of the Barabasi-Albert Model

- Characteristic Path Length
  - It can be shown (but it is difficult) that the BA model has a length proportional to *log(n)/log(log(n))*
  - Which is even shorter than in random networks
  - And which is often in accord with – but sometimes underestimates – experimental data
- Clustering
  - There are no analytical results available
  - Simulations shows that in scale-free networks the clustering decreases with the increases of the network order
  - As in random graph, although a bit less
  - This is not in accord with experimental data!



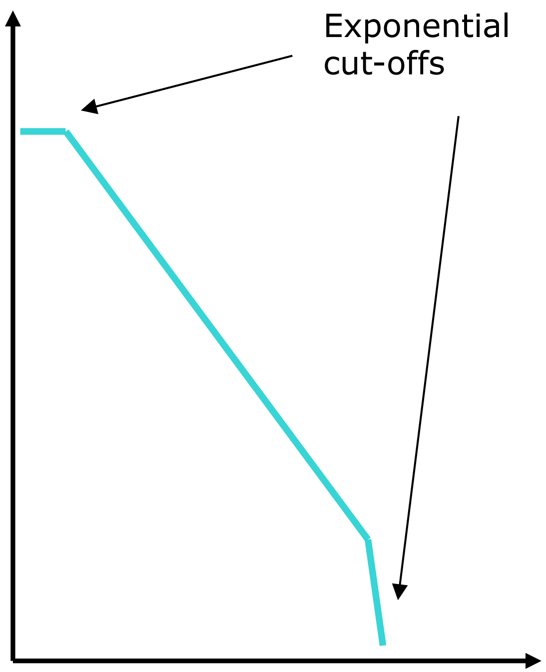


Problems of the Barabasi Albert Model (1)

- The BA model is a nice one, but is not fully satisfactory!
- The BA model does not give satisfactory answers with regard to clustering
  - While the small world model of Watts and Strogatz does!
  - So, there must be something wrong with the model..
- The BA model predicts a fixed exponent of 3 for the power law
  - However, real networks shows exponents between 1 and 3
  - So, there most be something wrong with the model

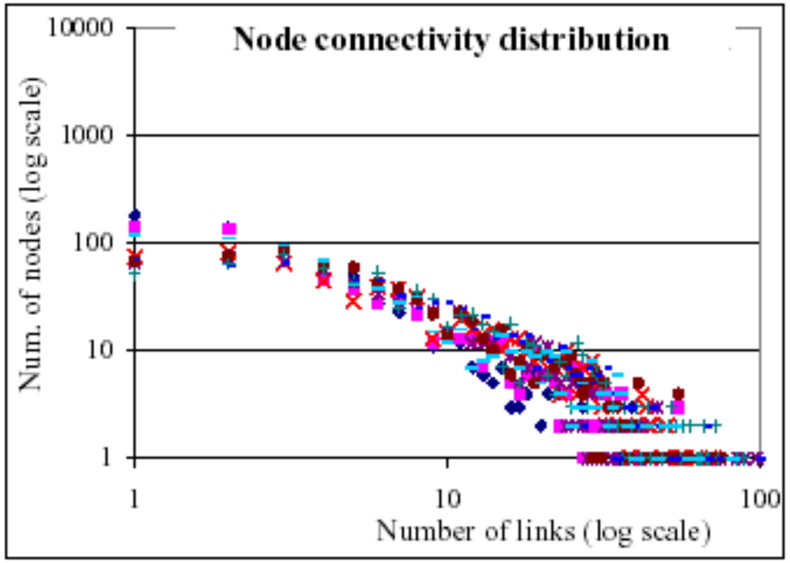
Problems of the Barabasi Albert Model (2)

- As an additional problem, is that real networks are not “completely” power law
  - They exhibit a so called ***exponential cut-off***
  - After having obeyed the power-law for a large amount of k
  - For very large k, the distribution suddenly becomes exponential
  - The same sometimes happen for
- In general
  - The distribution has still a “heavy tailed” is compared to standard exponential distribution
  - However, such tail is not infinite
- This can be explained because
  - The number of resources (i.e., of links) that an individual (i.e., a node) can sustain (i.e., can properly handled) is often limited
  - So, there can be no individual that can sustain any large number of resources
  - Viceversa, there could be a minimal amount of resources a node can have
- The Barabasi-Albert model not predict this



Exponential Cut-offs in Gnutella

- Gnutella is a network with exponential cut-offs
- That can be easily explained
  - A node cannot connect to the network without having a minimal number of connections
  - A node cannot sustain an excessive number of TCP connections



Variations on the Barabasi-Albert Model: Non-linear Preferential Attachments

- One can consider non-linear models for preferential attachment
  - E.g. $\Pi(k) \propto k^{\alpha}$
- However, it can be shown that these models destroy the power-law nature of the network

Variations on the Barabasi-Albert Model: Evolving Networks

- The problems of the BA Model may depend on the fact that networks not only grow but also evolve
  - The BA model does not account for evolutions following the growth
- Which may be indeed frequent in real networks, otherwise
  - Google would have never replaced Altavista
  - All new Routers in the Internet would be unimportant ones
  - A Scientist would have never the chance of becoming a highly-cited one

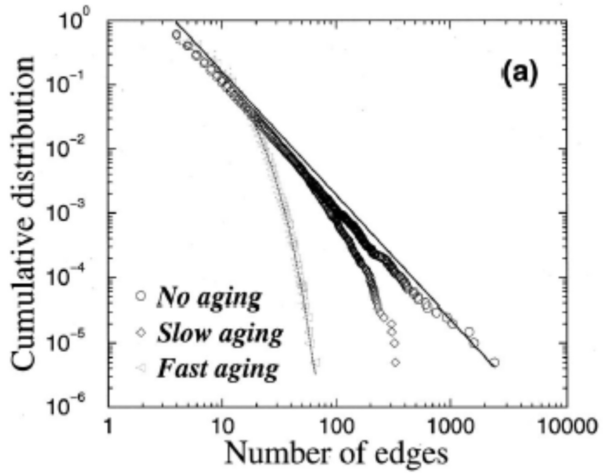
- A sound theory of evolving networks is still missing
  - Still, we can we start from the BA model and adapt it to somehow account for network evolution
  - And Obtain a bit more realistic model

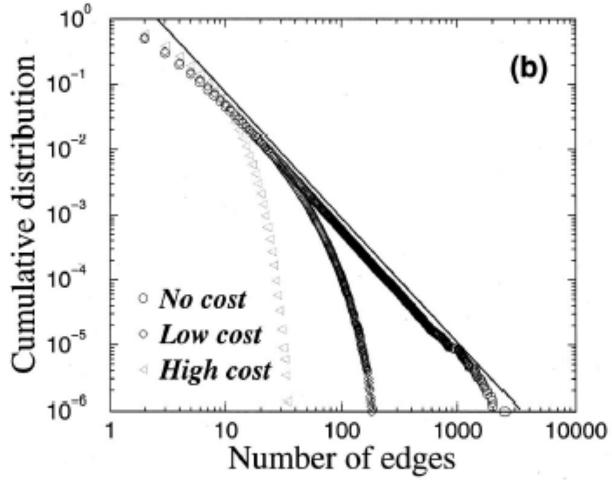
Variations on the Barabasi-Albert Model: Edges Re-Wiring

- By coupling the model for node additions
  - Adding new nodes at new time interval
- One can consider also mechanisms for edge re-wiring
  - E.g., adding some edges at each time interval
  - Some of these can be added randomly
  - Some of these can be added based on preferential attachment
- Then, it is possible to show (Albert and Barabasi, 2000)
  - That the network evolves as a power law with an exponent that can vary between 2 and infinity
  - This enables explaining the various exponents that are measured in real networks

Variations on the Barabasi-Albert Model: Aging and Cost

- One can consider that, in real networks (Amaral et al., 2000)
- Link cost
  - The cost of hosting new link increases with the number of links
  - E.g., for a Web site this implies adding more computational power, for a router this means buying a new powerful router
- Node Aging
  - The possibility of hosting new links decreased with the "age" of the node
  - E.g. nodes get tired or out-of-date
- These two models explain the "exponential cut-off" in power law networks





Variations on the Barabasi-Albert Model: Fitness

- One can consider that, in real networks
- Not all nodes are equal, but some nodes “fit” better specific network characteristics
  - E.g. Google has a more effective algorithm for pages indexing and ranking
  - A new scientific paper may be indeed a breakthrough
- In terms of preferential attachment, this implies that
  - The probability for a node of attracting links is proportional to some fitness parameter μ_i
  - See the formula below
- It can be shown that the fitness model for preferential attachment enables even very young nodes to attract a lot of links

$$\Pi(k_i) = \frac{\mu_i k_i}{\sum_j \mu_j k_j}$$

Summarizing

- The Barabasi-Albert model is very powerful to explain the structure of modern networks, but has some limitations
- With the proper extensions (re-wiring, node aging and link costs, fitness)
  - It can capture the structure of modern networks
  - The “rich get richer” phenomenon
  - As well as “the winner takes it all phenomena”
    - In the extreme case, when fitness and node re-wiring are allowed, it may happens that the network degenerates with a single node that attracts all link (monopolistic networks)
- Still, a proper unifying and sound model is missing

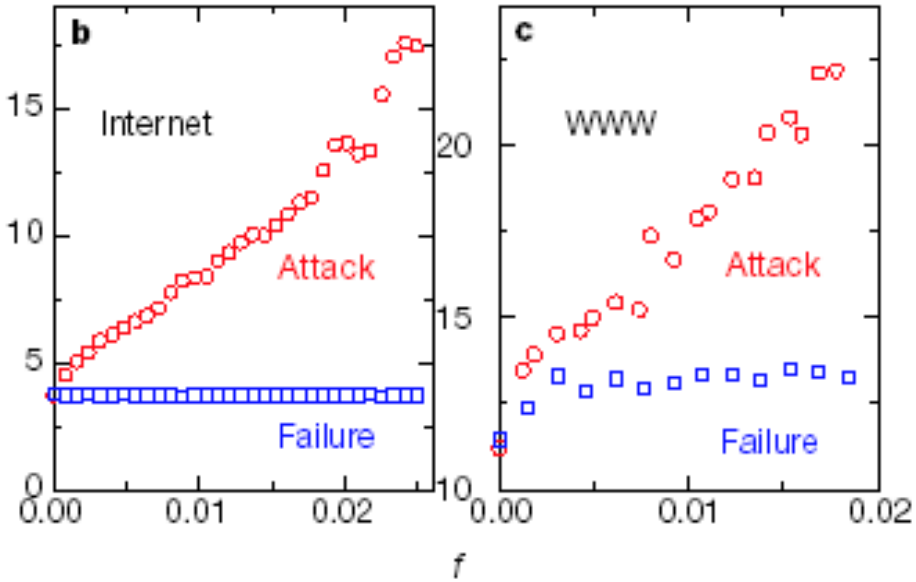
Part 3

- Properties of Scale Free Networks

Error Tolerance

- Scale free networks are very robust to errors
  - If nodes randomly "break" of disconnect to the network
  - The structure of the network, with high probability, will not be significantly affected by such errors
  - At least only a few small clusters of nodes will disconnect to the network
  - The average path length remains the same

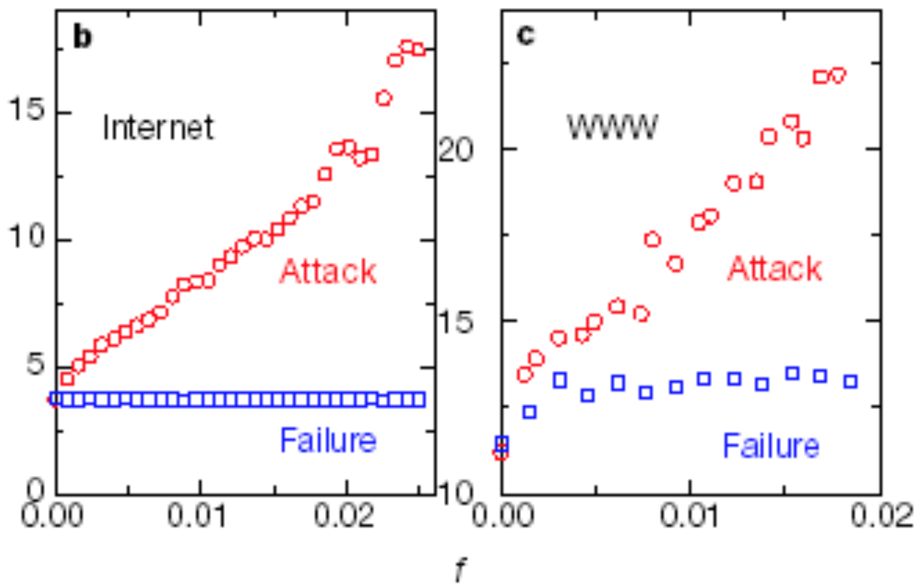
Characteristic Path Lenght



Attack Tolerance

- Scale free networks are very sensitive to targeted attacks
  - If the most connected nodes get deliberately chosen as targets of attacks
  - The average path length of the network grows very soon
  - It is very likely that the network will break soon into disconnected clusters
  - Although these independent clusters still preserves some internal connection

Characteristic Path Lenght



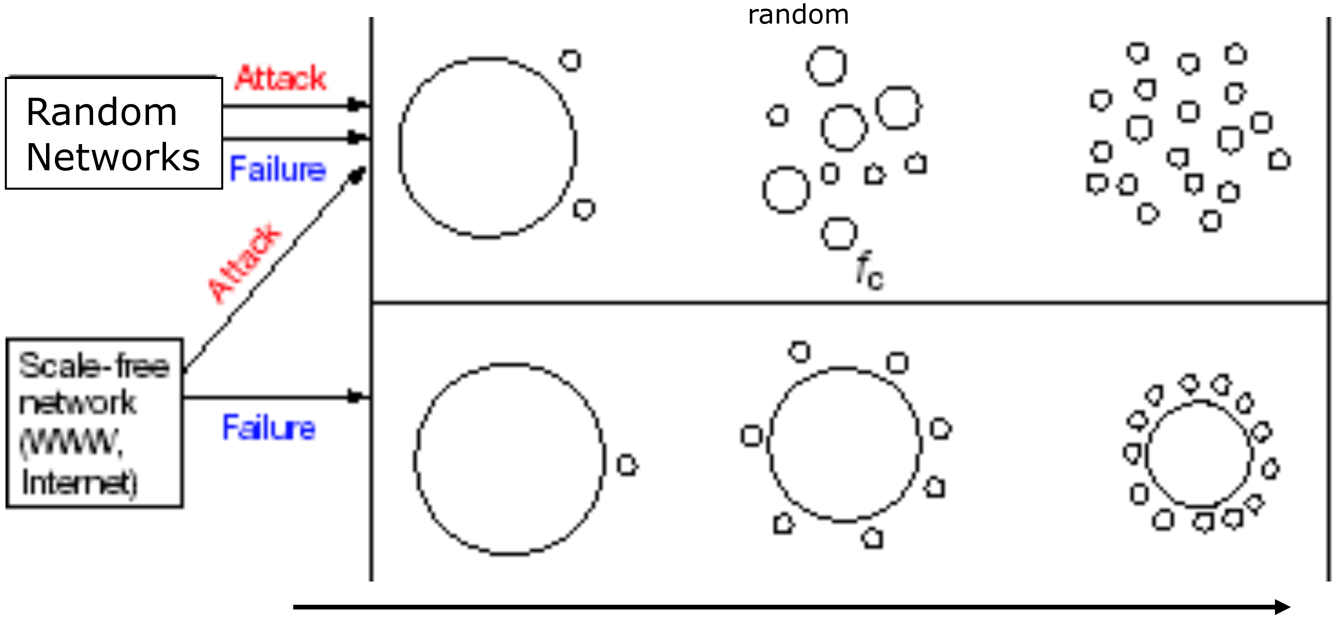
Error and Attack Tolerance: Random vs. Scale Free Networks

- Let us compare how these types of networks evolve in the presence of errors and attacks

For very limited errors/attacks, both networks preserve the connected structure

For increasing, but still very limited errors/attacks
(1) The random network break
(2) The scale free network breaks if the errors are targeted attacks!
(3) The scale free network preserve its structure if the errors are random

For relevant errors/attacks
(1) The random network break into very small clusters
(2) The scale free network do the same if the errors are targeted attacks!
(3) The scale free network preserve a notably connected structure if the errors are random



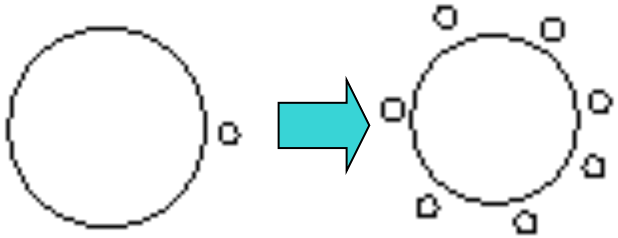
Increasing percentage of node errors/attacks

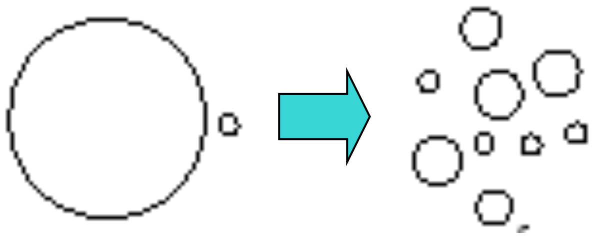
Epidemics and Percolation in Scale Free Networks (1)

- The percolation threshold p_c determines
  - the percentage of nodes that must be connected from a network to have the network be a single connected cluster
  - Or, the $(1-p_c)$ percenta ge of nodes that must be disconnected to have the network break into disconnected clusters
- Clearly, this is the same of saying
  - The percentage $(1-p_c)$ of nodes that must be immune to an infection for the infection not to become a "giant" one
- In fact
  - If the percentage $(1-p_c)$ of immune nodes are able to block the spreading of an infection
  - This implies that if these nodes were disconnected from the network, they would significantly break the network into a set of independent clusters
- This understood, what can be said about epidemics in scale free networks?

Epidemics and Percolation in Scale Free Networks (2)

- Given that a scale-free network
  - In the presence of even a large amount of random errors
  - Does not significantly break into clusters (see Figure 2 slides before)
- This implies that the percolation threshold p_c in scale free network is practically zero
  - There is no way to stop infections in random nodes even when a large percentage of the population is immune to them!!!
- On the other hand
  - If we are able to make immune the mostly connected nodes
  - Breaking the network into independent clusters
  - That is, if the immune nodes are not selected at random by in the most effective way
- Then, in this case, we can stop infections in a very effective way!





Implications for Distributed Systems: Internet Viruses and Routers' Faults

- There is practically no way to break the spread of Internet viruses
  - But by immunizing the most relevant "hub" routers
- The structure of the Internet is very robust in the presence of router faults
  - Several routers can fails, and they do everyday, without causing significant partitionings of the network
- At the same time
  - If very important "hub" routers fails, the whole network can suddenly become disconnected
  - E.g., the destroying of World-Trade-Center routers – acting as main hubs for Europe-America connections – on September 11

Implications for Distributed Systems: Web Visibility

- How can we make our Web site a success?
  - We must make sure that it is connected (incoming links especially) from a relevant number of important sites
  - Search engines, clearly, but also all our clients
  - This will increase the probability of it becoming more and more visible
- We must make sure that it has “fitness”
  - What added value does it carry?
  - Can such added value increase its probability of preferential attachment?
- However, we must always consider that random processes still play an important role

Implications for Everyday Systems: Scale Free Networks and Trends

- Who decide what is in and what is “out” in music, fashion, etc.?
  - How can an industry have its products become “in”?
- Industries spend a lot of money in trying to influence the market
  - A lot of commercial advertising, a lot of “free trials”, etc.
  - Still, many new products fail and never have market success!
- Recently, a few innovative industries have tried to study the structure of social network
  - And have understood that to launch a new product is important to identify the “hubs” of the social network
  - And have this hubs act as the engine for the launch of the product
- To this end, their commercial strategy consider
  - Recruiting and paying people of the social layer they want to influence
  - Send this people to discos, pubs, etc.
  - And identify the “hubs” (i.e., the smart guys that in the pub knows everybody, is friendly and has a lot of women,
  - After which, paying such identified hubs to support the product (e.g., wearing a new pair of shoes)
- Reebok did this by giving free shoes in suburbia basket camps in US
  - Thus conquering the afro-american market

Implications for Everyday Systems: Scale Free Networks and Terrorism

- The network of terrorism is growing
  - And it is a social network with a scale free structure
- How can we destroy such network?
  - Getting unimportant nodes will not significantly affect the network
  - Getting the right nodes, i.e., the hubs (as Bin Laden) is extremely important
  - But it may be very difficult to identify and get the hubs
  - In any case, even if we get the right nodes, other connected clusters will remains that will likely act in any case
- As far as breaking the information flow among terrorists
  - This is very difficult because of the very low percolation threshold

Conclusions and Open Issues

- In the modern "complex networks" theory
  - Neither small world nor small free networks captures all essential properties of real networks (and of real systems)
  - However, both systems capture some interesting properties
- In the future, we expect
  - More theories to emerge
  - And more analysis on the dynamic properties of these types of network (i.e., of what happens when there are processes running over them) to be performed
- This will be of great help to
  - Better predict and engineer the networks themselves and the distributed application that have to run over them
  - Apply phenomena of self-organization in nature (mostly occurring in space) to complex networks in a reliable and predictable ways